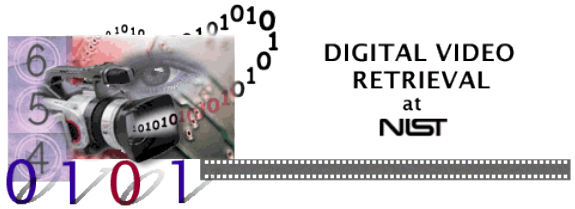
Multimedia Event Detection Task

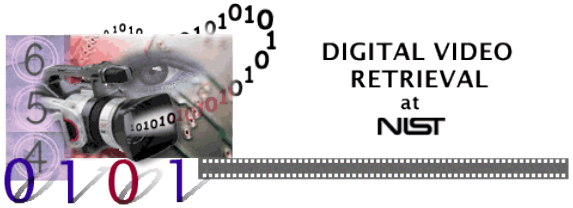
Time		Presentation
9:20	– 9:40	Task Overview (NIST)
9:40	– 10:00	Access to Audiovisual Media (AXES)
10:00	– 10:20	SRI International; Sarnoff Corporation (Aurora)
10:20	– 10:40	Break in the NIST West Square Cafeteria
10:40	– 11:00	Kitware Inc (GENIE)
11:00	– 11:20	Tokyo Institute of Technology; Canon Corporation (TokyoTechCanon)
11:20	– 11:40	SRI Internations (SESAME)
11:40	– 12:10	Discussion



2012 TRECVID Workshop Multimedia Event Detection Task

Jonathan Fiscus
National Institute of Standards and Technology (NIST)

Martial Michel
Systems Plus Inc.



Talk Outline

- MED Task Overview
- HAVIC Data Resources
- The 2012 MED Results
- Summary and What's Next

MED Task Definition

Given an event specified by an **event kit**, search multimedia recordings for the event:

1. determine a hard decision confidence threshold prior to search time,
2. assign a confidence score to each clip in the collection,
3. measure Content Description build time, and
4. measure the Event Agent execution time

An MED Event is

- complex activity occurring at a specific place and time;
- involves people interacting with other people and/or objects;
- consists of a number of human actions, processes, and activities that are loosely or tightly organized and that have significant temporal and semantic relationships to the overarching activity;
- is directly observable.

Rock Climbing Event Kit Text

Definition:

One or more people climb up or across rock formations or artificial rock walls.

Explication:

Rock climbing is a physically intense activity, where the goal is to reach the top or endpoint of a pre-defined route on a rock formation or artificial rock wall by finding a grip on the surface using hands and feet, and then pulling up using their arm and leg strength. …

Evidential Description:

- scene: outdoors in natural setting, indoors in rock climbing gym, or outdoors on a specially …
- objects/people: carabiners, rope, helmet, harness, rock formation, artificial rock wall, climbers
- activities: hooking rope to harness, moving hands and feet along side of rock face, grabbing rock …….
- audio: carabiners clinking, climbers making comments on the difficulty of the climb, onlookers cheering on …

Illustrative Examples

- Positive instances of the event
- Clips "Related" to the event

Positive Rock Climbing Video Example

MED Evaluation Conditions

- MED Tasks
  - **Pre-Specified Event (PS)** – MED metadata generation optimized with knowledge of events
  - **Ad-Hoc Event (AH)** – MED metadata generation complete before events revealed
- Event Agent Generation (EAG) Processing Types
  - **Automatic EAG** – No human interaction to build the event agent
  - **Semi-Automatic EAG** – Human guidance of event agent building
- Events Processes
  - **MEDFull** – Processing 20 PS event, 5 AH events
  - **MEDPart** – Processing a subset of the events
- Event Training Condition
  - **EKFull** – Use the event kit text and all supplied positive, near_miss, and related exemplars
  - **EK10Ex** – Use a 10-positive and 10-related clip subset (20 total) of EKFulll
- Required Condition
  - PS, EKFull

The TRECVID MED 2012 Events

Pre-Specified Events

MED '11 Events

Changing a vehicle tire
Getting a vehicle unstuck
Grooming an animal
Making a sandwich
Parkour
Repairing an appliance
Working on a sewing project
Birthday party
Flash mob gathering
Parade

New Events

Attempting a bike trick
Cleaning an appliance
Dog show
Giving directions to a location
Marriage proposal
Renovating a home
Rock climbing
Town hall meeting
Winning a race without a vehicle
Working on a metal crafts project

Ad Hoc Events

New Events

Doing homework or studying
Hide and seek
Hiking
Installing flooring
Writing text

17 MED 2012 Finishers and Number of Runs

Team	Ad-Hoc		Pre-Specified		Organization
	EK10X	EKFull	EK10X	EKFull	
AXES		4		4	Dublin City Univ., Univ. Twente, Oxford Univ., Katholieke University Leuven, Technicolor, Erasmus Univ. Rotterdam, Fraunhofer Gesellschaft, Cassidian, BBC, Deusche-Welle, Netherlands Inst. for Sound and Vision, ERCIM
BBNVISER*	2	1	1	2	Raytheon BBN Technologies, UMD, Columbia, UCF Team
CERTH-ITI*				4	Informatics and Telematics Inst., Centre for Research and Tech.
CMU*	1	1	1	3	Carnegie Mellon University
DCU-iAD-CLARITY*		1		1	Dublin City University, IAD
ECNU				3	Institute of Computer Applications, East China Normal University
Genie*	1	2	1	3	Kitware Inc.
IBMCU*	1	1	1	3	IBM T. J. Watson Research Center
MediaMill*	2	1		3	University of Amsterdam
NII*				4	National Institute of Informatics
NTT-NII		2		4	NTT Communication Science Lab., Nat. Inst. of Informatics
OPU		1		1	Osaka Prefecture University
Sesame*	1	2	1	3	SRI International SESAME
SRIAURORA*	1	1	1	3	SRI International Sarnoff Aurora
TokyoTechCanon*		2		4	Tokyo Institute of Technology and Canon
UEC		1		1	University of Electro-Communications
VIREO				3	City University of Hong Kong
	9	**20**	**6**	**49**	

* MED '11 Participants

Data Collection & Annotation

- Team of 50 data scouts at the Linguistic Data Consortium
  - In-person training, regular team meetings, work remotely
- Custom GUI to search web for appropriate videos, then annotate their properties
- Two guiding annotation principles, plus corollary
  - **Sufficient Evidence Rule:** Video must contain sufficient evidence to decide that an event has occurred
  - **Reasonable Viewer Rule:** If according to a reasonable interpretation of the video the event must have occurred, then the clip is a positive instance of that event
  - **Corollary**: Not necessary for full process to be shown
- Scouts encouraged to seek out interesting, varied clips

Annotation and Preparation of Candidate Videos

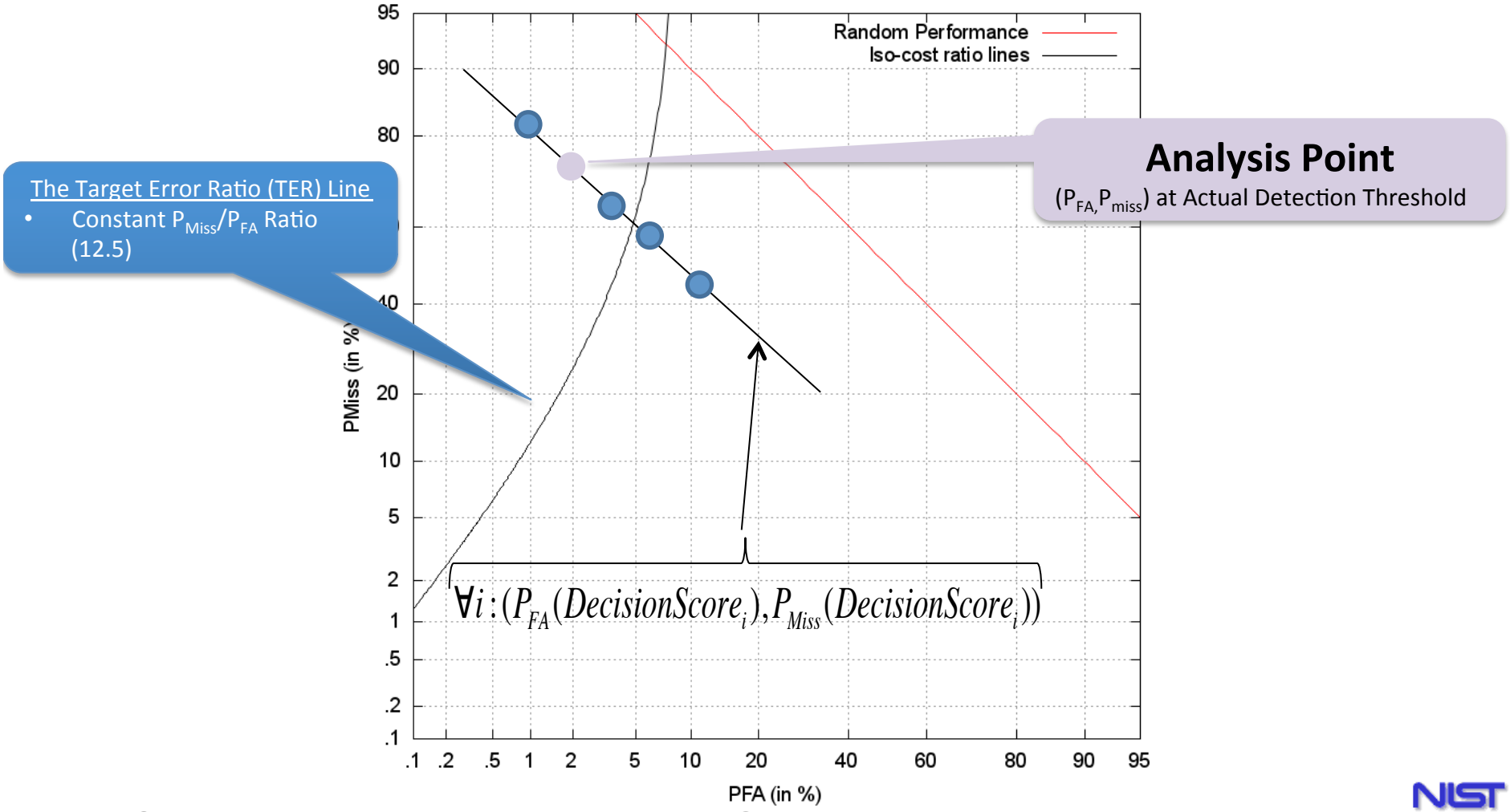
- For each candidate video, scouts are required to
  - Watch clip in its entirety
  - Determine and verify the download URL
  - Screen for sensitive PII, objectionable content
- Collection strategies
  - Event specific: label with event status (positive, near miss, background)
  - Background clips: collected without regard to an event
- Downloaded videos processed to standardize data format and encoding
  - MPEG-4, h.264 video encoding, aac audio encoding
  - Original video resolution and audio/video bitrates retained

HAVIC Data Resources

		Video clips	Video duration
MED '12 Training	MED '10	3,468	114 hours
	MED '11 DEV	10,403	324 hours
	MED '11 Eval	32,061	991 hours
	Transcription	1,498	45 hours
Progress Test Collection (Used for MED '12-15)		98,117	3,722 hours
Total		144,049	5,151 hours

Decision Error Tradeoff (DET) Curves

$Prob_{Miss}$ vs. $Prob_{FA}$

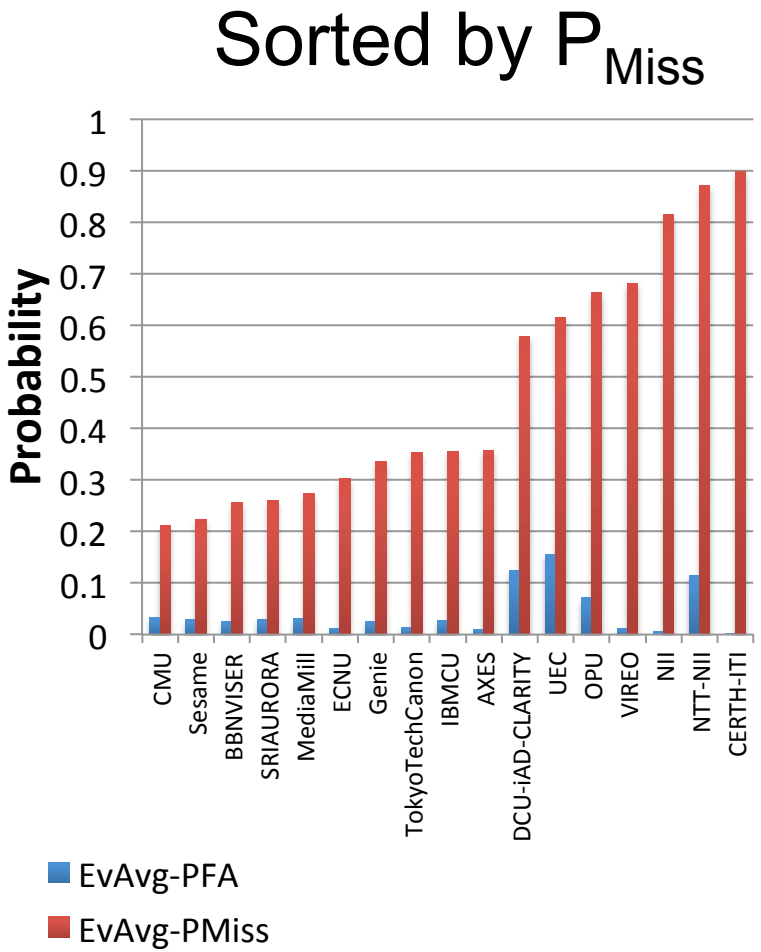


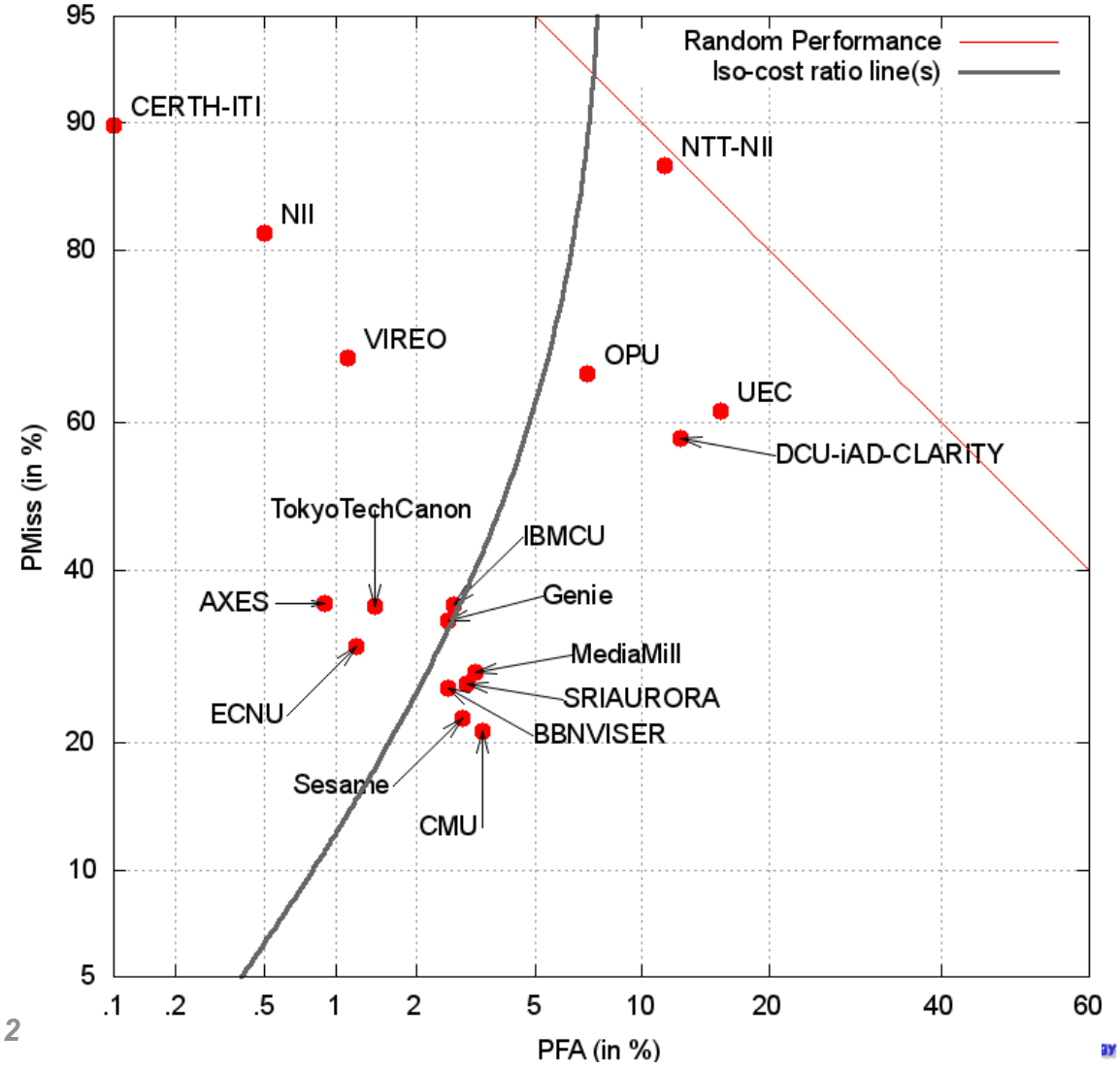
The Target Error Ratio (TER) Line
- Constant P_{Miss}/P_{FA} Ratio (12.5)

Analysis Point
(P_{FA},P_{miss}) at Actual Detection Threshold

Primary, Pre-Specified Event Systems

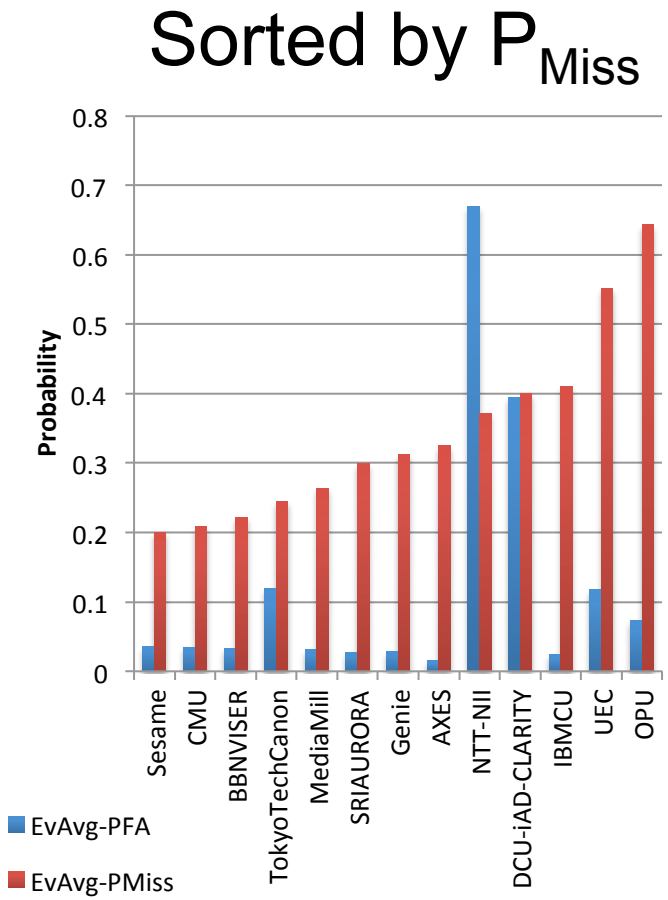
Event-Averaged, P_{Miss} and P_{FA} at Actual Decision Threshold

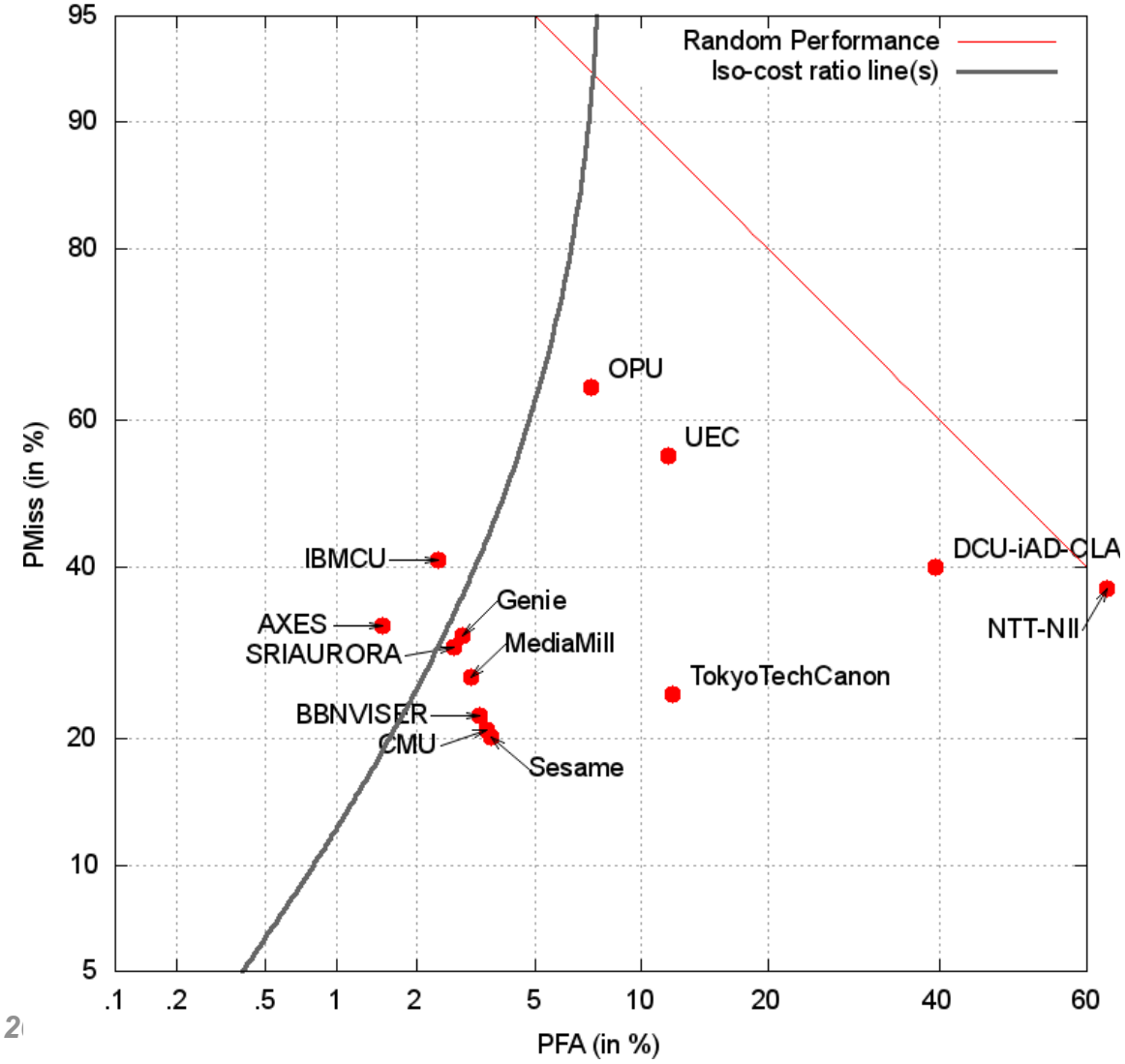




Primary, Ad-Hoc Event Systems

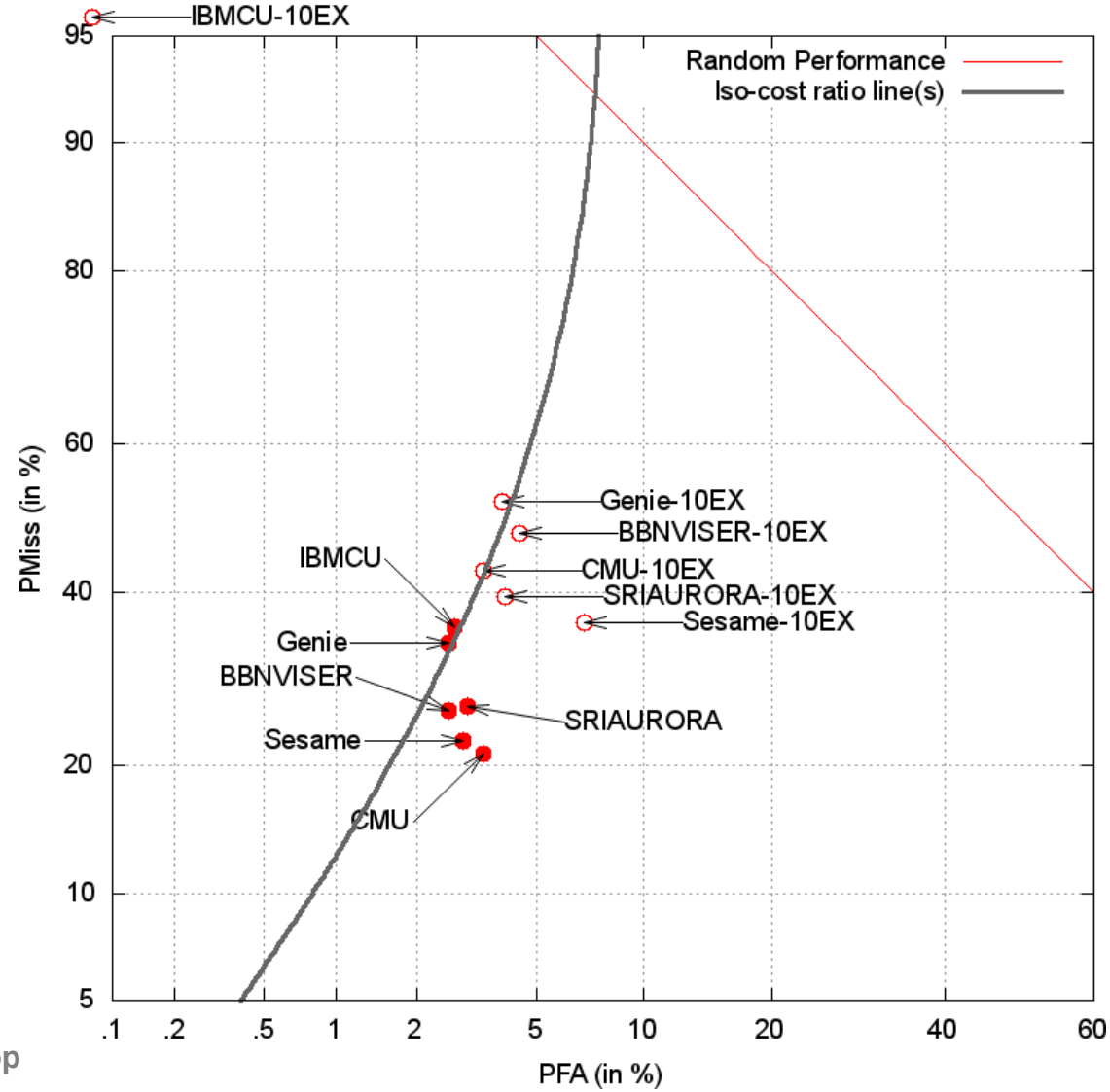
Event-Averaged, P_{Miss} and P_{FA} at Actual Decision Threshold





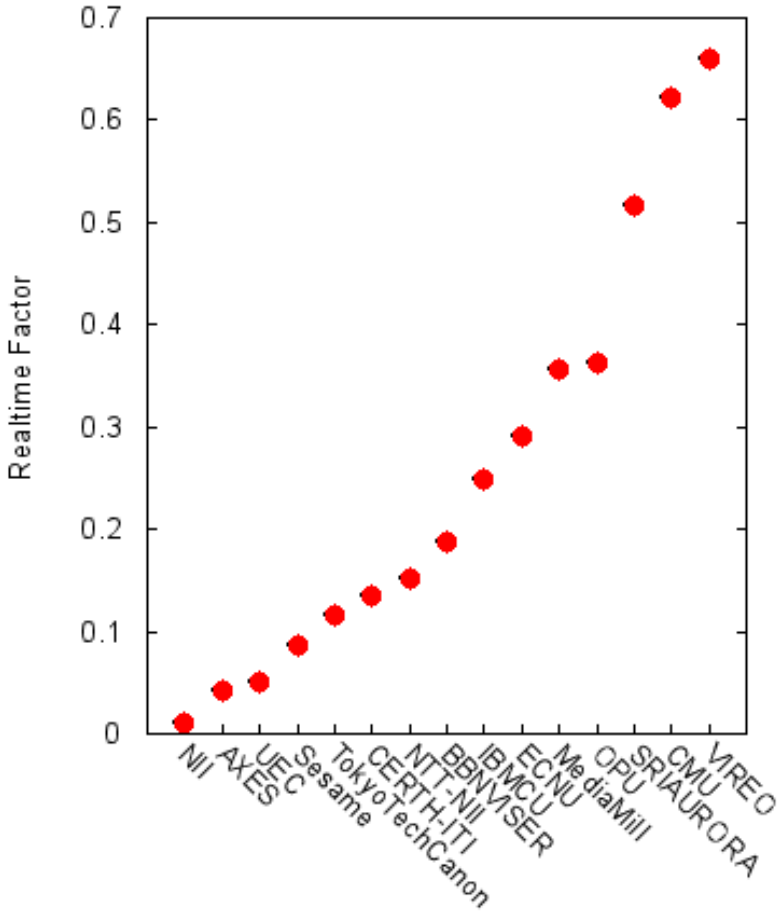
Pre-Specified Event Systems: EKFull vs. EK10Ex

Event-Averaged, P_{Miss} and P_{FA} at Actual Decision Threshold

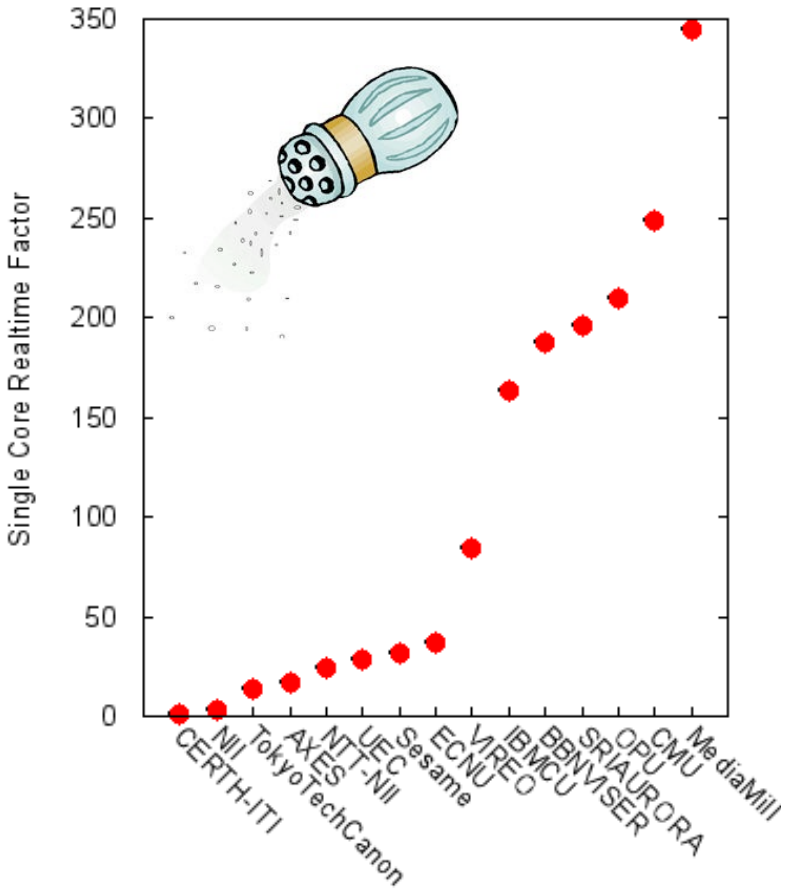


Metadata Generation Speed
Primary Pre-Specified Event Systems

Realtime Factors

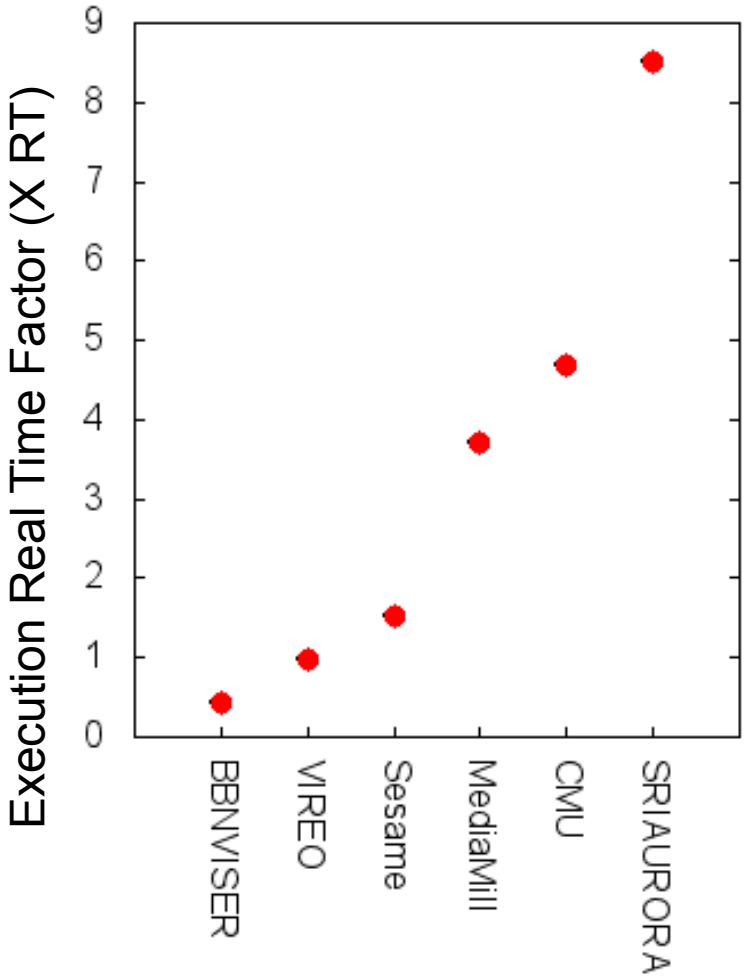


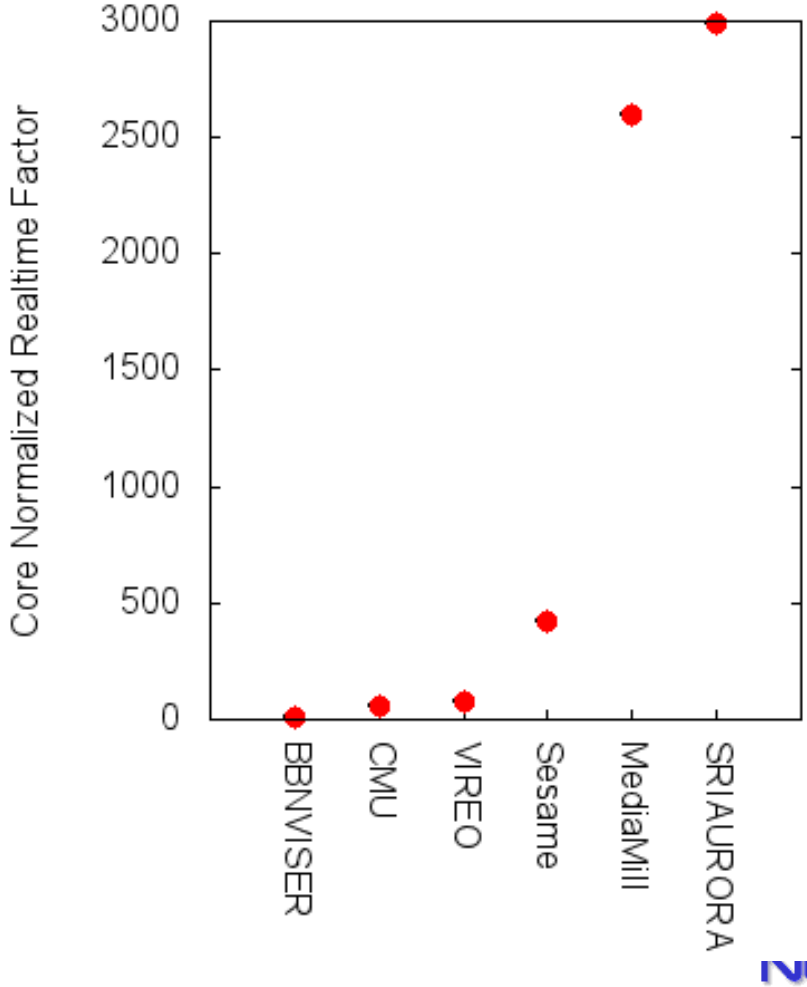
Single Core Realtime Factors



Event Agent Execution Speed

Pre-Specified Event Systems





Summary

- 17 participants processed 3,722 hours of video searching for 20 Pre-Specified events
  - Reporting results on the Progress Set while protecting the data set's properties was challenging
    - Reporting threshold performance would give too much insight into the data set
  - Core-normalized runtimes should be viewed as approximate
- 13 participants participated in the Ad-Hoc Event Pilot
  - Encouraging results for the Ad-Hoc Events
- What's next?
  - The Progress set should be not accessible until next summer
    - Remove LDC disk drive from your system
    - Either: delete video and models, turn off disk drive(s), or remove drive(s)
  - Proposed task changes (NIST will present during the discussion period)
  - Next evaluation cycle will:
    - Retest the 25 MED '12 events
    - Add 10 new Pre-specified events (Spring)
    - Add 10 new AdHoc events (Fall)

Questions?

Backup Slides

Proposed 2013 MED Task Changes

- Discontinue event kit texts
- Define Negative clips as a unique set of file IDs
- Test MED systems on event exemplars
- Combine MER and MED into a single system output per clip